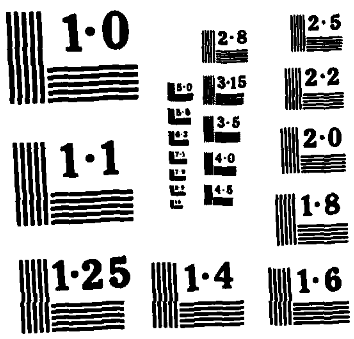
AD-A157 681 PROPERTIES OF A REPRESENTATION OF A BASIS FOR THE NULL SPACE(U) STANFORD UNIV CA SYSTEMS OPTIMIZATION LAB P E GILL ET AL. FEB 85 SOL-85-1 N00014-75-C-0267 1/1

UNCLASSIFIED F/G 12/1 NL

END
FILMED
DTIC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

ARO 21592.5-MA

2

Systems
Optimization
Laboratory

PROPERTIES OF A REPRESENTATION OF A BASIS FOR THE NULL SPACE

by

Philip E. Gill†, Walter Murray†, Michael A. Saunders†,
G. W. Stewart* and Margaret H. Wright†

TECHNICAL REPORT SOL 85-1

FEBRUARY 1985

Department of Operations Research
Stanford University
Stanford, CA 94305

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

PROPERTIES OF A REPRESENTATION OF A BASIS FOR THE NULL SPACE

by

Philip E. Gill[†], Walter Murray[†], Michael A. Saunders[†],
G. W. Stewart[*] and Margaret H. Wright[†]

TECHNICAL REPORT SOL 85-1

FEBRUARY 1985

Research and reproduction of this report were partially supported by National Science Foundation Grants MCS-7926009 and ECS-8312142; U.S. Department of Energy Contract DE-AM03-76F00326, PA# DE-AT03-76ER72018; Office of Naval Research Contract N00014-75-C-0267; and U.S. Army Research Office Contract DAAG29-84-K-0156 at Stanford, and Air Force Office of Scientific Research Contract AFOSR-82-0078 at the University of Maryland, College Park.

[†]Department of Operations Research, Stanford University, Stanford, California 94305.

[*]Department of Computer Science, University of Maryland, College Park, Maryland 20742.

A

PROPERTIES OF A REPRESENTATION OF A BASIS FOR THE NULL SPACE

by

Philip E. Gill[†], Walter Murray[†], Michael A. Saunders[†], G. W. Stewart[‡] and Margaret H. Wright[†]

ABSTRACT

Sub 1

Given a rectangular matrix $A(x)$ that depends on the independent variables x, many constrained optimization methods involve computations with $Z(x)$, a matrix whose columns form a basis for the null space of $A^T(x)$. When A is evaluated at a given point, it is well known that a suitable Z (satisfying $A^TZ = 0$) can be obtained from standard matrix factorizations. However, Coleman and Sorensen have recently shown that standard orthogonal factorization methods may produce orthogonal bases that do not vary continuously with x; they also suggest several techniques for adapting these schemes so as to ensure continuity of Z in the neighborhood of a given point.

This paper is an extension of an earlier note that defines the procedure for computing Z. Here, ~~we~~ The authors first describe how Z can be obtained by *updating* an explicit QR factorization with Householder transformations. The properties of this representation of Z with respect to perturbations in A are discussed, including explicit bounds on the change in Z. ~~We~~ They then introduce *regularized* Householder transformations, and show that their use implies continuity of the full matrix Q. The convergence of Z and Q under appropriate assumptions is then proved. Finally, ~~we~~ it is indicate~~s~~ why the chosen form of Z is convenient in certain methods for nonlinearly constrained optimization.

1. Introduction

Given an $n \times m$ matrix A of rank m ($m < n$), many constrained optimization methods involve computations with a (non-unique) matrix Z whose $n - m$ columns form a basis for the null space of A^T (i.e., such that $A^T Z = 0$). Typically, A represents the Jacobian of a set of constraints, and the elements of A are smooth functions of an independent variable x ($x \in \Re^n$). Attention has recently been focussed on the *continuity properties* of the associated Z, which turn out to be crucial in proving local convergence for certain methods. For example, in Coleman and Conn (1982, 1984), an essential assumption is that small changes in x lead to small changes in Z.

It is well known that certain factorizations of A provide stable and efficient means for computing Z. For example, given the QR factorization of A,

$$A = Q \begin{pmatrix} R \\ 0 \end{pmatrix}, \tag{1}$$

where Q is an $n \times n$ orthogonal matrix, and R is an $m \times m$ non-singular upper-triangular matrix, Q may be partitioned as

$$Q = (\overbrace{Y}^{m} \quad \overbrace{Z}^{n-m}). \tag{2}$$

Coleman and Sorensen (1984) observed that the standard method of computing the QR factorization through Householder matrices may not provide a continuous representation of $Z(x)$. They proposed several alternative strategies for ensuring a continuous Z, based on removing the discontinuity associated with the sign that defines each Householder transformation. Gill *et al.* (1983) present an updating technique that provides a continuous representation of Z. Byrd and Schnabel (1984) note that inherent discontinuities exist if Z is defined as a function of x, except in certain special cases.

This paper is an extension of Gill *et al.* (1983). The matrix Z is a submatrix of an explicit orthogonal matrix Q, which is obtained by *updating* the QR factorization. In Section 2, we summarize the procedure for computing Z and Q. In Section 3 we give explicit bounds for the change in Z resulting from perturbations in x, and show that Z is continuous in the neighborhood of a point where A has full rank. We then introduce the class of *regularized* Householder transformations, analyze the effect of perturbations in x on the full matrix Q, and give a similar proof of continuity. In Section 4, we prove that Z approaches a limit when computed at a sequence of points $\{x_k\}$ converging sufficiently fast to a suitable point x^* (and similarly for Q when regularized Householder transformations are used). Numerical examples are given in Section 5 to illustrate some of the results. Finally, in Section 6 we discuss the chosen representation for Z in the context of algorithms for constrained optimization.

2. Representation and computation of Z

It is essential to distinguish between the theoretical definition of Z as a matrix whose columns have specified properties, and its realization as a data structure with which computations are performed. Although a "matrix" may be used as a pedagogical convenience in describing an algorithm, it will not necessarily be represented as an explicit two-dimensional data structure within an implementation. For example, the standard Householder method for computing the QR factorization (1) (see, e.g., Stewart, 1973) results in a special sequence of m Householder transformations that are stored in *compact form* (each represented by a vector). This *implicit form of* Q is acceptable in many contexts — in particular, most optimization algorithms do not require the elements of Z, but rather only the ability to compute products involving Z and its transpose. With an implicit Q, operations with Z are performed by applying the sequence of transformations (not by explicit matrix multiplication).

In contrast, the procedure to be described obtains Z from a QR factorization in which Q is stored *explicitly*. We assume that $A(x)$ is the Jacobian of a mixture of linear and nonlinear constraints. Accordingly, let the m columns of $A(x)$ be partitioned into two groups: the first m_L columns (denoted by A_L, and termed the constant columns) are independent of x, and correspond to the gradients of *linear* constraints; the last m_N columns (denoted by $A_N(x)$, and termed the variable columns) vary with x, and correspond to the gradients of *nonlinear* constraints. Thus, A and R in (1) have the forms

$$A = (A_L \quad A_N) \qquad \text{and} \qquad R = \begin{pmatrix} R_L & T \\ 0 & R_N \end{pmatrix},$$

where R_L and R_N are upper-triangular. *The factorization (1) of A is assumed to be available, with Q stored explicitly.* Note that

$$A_L = Q\begin{pmatrix} R_L \\ 0 \end{pmatrix}. \tag{3}$$

Now consider a different matrix $\bar{A}$, given by

$$\bar{A} = (A_L \quad \bar{A}_N). \tag{4}$$

It follows from (3) that

$$\bar{A} = Q\begin{pmatrix} R_L & \bar{T} \\ 0 & V \end{pmatrix},$$

where V is $(n - m_L) \times m_N$. Thus, Q triangularizes the first m_L columns of $\bar{A}$, and the QR factorization of $\bar{A}$ can be obtained by triangularizing V. In order to perform this computation, we apply m_N *updates* to the QR factorization (3) of A_L while the columns of $\bar{A}_N$ are added one at a time. The crucial point about this approach is that *the factorization is updated after each column is added.* Thus, after the i-th column of $\bar{A}_N$ has been processed, an explicit orthogonal matrix is available that triangularizes columns 1 through $m_L + i$ of $\bar{A}$.

To describe the work associated with each update, we consider an $n \times (j-1)$ matrix $\tilde{C}$ whose QR factorization is given by

$$\tilde{C} = \tilde{Q}\begin{pmatrix} \tilde{R} \\ 0 \end{pmatrix}. \tag{5}$$

Let $\hat{C} = (\tilde{C} \quad c)$ for some vector c. Then, from (5),

$$\tilde{Q}^T\hat{C} = (\tilde{Q}^T\tilde{C} \quad Q^Tc) = \begin{pmatrix} \tilde{R} & t \\ 0 & v \end{pmatrix}. \tag{6}$$

We now construct the Householder transformation H_j that leaves t unaltered and annihilates all but the first element of v. If

$$v = \begin{pmatrix} \nu \\ \tilde{v} \end{pmatrix},$$

where ν is the j-th diagonal element of $\tilde{Q}^T\hat{C}$, then

$$H_j(v) = I - \frac{1}{\beta}uu^T,$$

where

$$u = (\overbrace{0,\ldots,0}^{j-1}, \nu + \operatorname{sign}(\nu)\|v\|, \tilde{v}^T), \quad \beta = \frac{1}{2}\|u\|^2. \tag{7}$$

(Unless otherwise stated, $\|\cdot\|$ denotes the Euclidean vector norm or the induced matrix norm.) We then have

$$\hat{C} = \tilde{Q}H_j \begin{pmatrix} \tilde{R} & t \\ 0 & \rho \\ 0 & 0 \end{pmatrix} = \hat{Q} \begin{pmatrix} \hat{R} \\ 0 \end{pmatrix},$$

where $\rho = -\operatorname{sign}(\nu)\|v\|$ and

$$\hat{Q} = \tilde{Q}H_k. \tag{8}$$

The following should be noted: only columns j through n of $\tilde{Q}$ are altered by H_k; H_j depends only on v (not on t); and $H_j(\alpha v) = H_j(v)$, where α is any non-zero scalar.

In order to obtain the factors of $\bar{A}$ (4), the above procedure is repeated m_N times, beginning with $\tilde{C} = A_L$, $\tilde{R} = R_L$, and $\tilde{Q} = Q$; each column of $\bar{A}_N$ then takes the role of c in (6). With this approach, a "current" orthogonal matrix (which, for simplicity, we shall denote by $\tilde{Q}$) is always available after each column is added. Because each Householder transformation is applied to $\tilde{Q}$ before the next transformation is constructed, $\tilde{Q}$ represents *all* previous transformations. By applying $\hat{Q}$ to the new column as the first step in each update, the effect is to apply the entire sequence of Householder transformations. Therefore, each Householder matrix is *not* applied to the remaining (untransformed) columns of $\bar{A}_N$, in contrast to the standard Householder procedure.

Completion of a single update involves three steps: (i) formation of $\tilde{Q}^Tc$ to obtain v, (ii) definition of u, and (iii) application of the Householder transformation to $\tilde{Q}$. The desired factorization of $\bar{A}$ requires m_N updates, and the final $\bar{Q}$ satisfies

$$\bar{Q}^T = H_m \cdots H_1 Q^T. \tag{9}$$

The total work required to obtain $\bar{Q}$ and $\bar{R}$ is of the order of $2nm_N(n - m_L) + nm_N(n - m_N)$ operations.

3. Perturbation analysis of Z and Q

In this section we analyze the effect of the procedure of Section 2 when applied at points "near" a point $\hat{x}$ where $A(\hat{x})$ has full rank. Roughly speaking, the desired continuity properties involve showing that small changes in x lead to small changes in Z. The proof is complicated by the fact that the class of Householder transformations does not include the identity matrix or any matrix near it. Therefore, considering (9), the full matrix Q is *not* continuous. However, it turns out that small changes in x do lead to small changes *in the columns of* Z. Explicit bounds on the perturbation in Z are derived in Section 3.1.

Although this result is satisfactory for methods in which only Z is required, continuity of *all* of Q is useful in other contexts — for example, when an explicit representation of the range space is used in an update. To extend the continuity result to all of Q, in Section 3.2 we introduce the class of regularized Householder matrices (which does contain the identity), and show that bounds similar to those for Z in the standard case can be obtained for all of Q when updates are performed with regularized Householder transformations.

3.1. Perturbation in Z. For simplicity, in this section we assume that $m_L = 0$, i.e., that all columns of A are variable; the analysis can be applied in a straightforward manner when constant columns are present. Given any $\epsilon_x > 0$ and the associated neighborhood of points $\hat{x} + \delta x$, where

$$\|\delta x\| < \epsilon_x, \tag{10}$$

we analyze the computation of $Z(\hat{x} + \delta x)$ from $Z(\hat{x})$ using the procedure of Section 2.

Let A denote $A(\hat{x})$, and $\bar{A}$ denote $A(\hat{x} + \delta x)$, with a similar convention for Q and Z. The QR factorization (1) of A is assumed to be given. Since A is a twice-continuously differentiable function of x, given any $\epsilon > 0$, there exists ϵ_x such that (10) implies

$$\bar{A} = A + \delta A, \quad \text{where} \quad \|\delta A\| \le \epsilon. \tag{11}$$

The existing Q^T "almost" triangularizes $\bar{A}$, i.e.,

$$Q^T\bar{A} = \tilde{R} = \begin{pmatrix} R \\ 0 \end{pmatrix} + E, \tag{12}$$

where $\|E\| \le \epsilon$. In order to triangularize $\tilde{R}$, the procedure of Section 2 constructs a special sequence of m Householder transformations $\{H_1, \ldots H_m\}$. Thus, we have

$$\bar{A} = \bar{Q}\begin{pmatrix} \bar{R} \\ 0 \end{pmatrix}, \quad \text{with} \quad \bar{Q}^T = H_m \cdots H_1 Q^T. \tag{13}$$

The jth transformation H_j is constructed so that its application to a vector v does not alter components 1 through $j-1$, and annihilates components $j+1$ through n.

The matrix $\bar{Z}$ corresponding to $\bar{A}$ comprises the last $n-m$ columns of $\bar{Q}$. To examine the changes in this part of Q, we introduce a sequence of diagonal matrices $\{D_j\}$, $j = 1, \ldots, m$, with

$$D_j = \mathrm{diag}(\overbrace{0, \ldots, 0}^{j}, \overbrace{1, \ldots, 1}^{n-j}).$$

Then, from (2),

$$QD_m = (0 \quad Z).$$

Since

$$\bar{Q}^T - Q^T = (H_m \cdots H_1 - I)Q^T$$

and $\|Q\| = 1$, we obtain the following bound for the change in Z:

$$\|\bar{Z} - Z\| \le \|D_m(I - H_m \cdots H_1)\|. \tag{14}$$

To derive a bound for the right-hand side of (14), observe that the special structure of H_j implies that

$$D_j H_j = D_j H_j D_{j-1}. \tag{15}$$

Using (15), the fact that $\|D_j H_j\| \le 1$ and the identity

$$D_j(I - H_j \cdots H_1) \equiv D_j(I - H_j) + D_j H_j (I - H_{j-1} \cdots H_1),$$

we obtain

$$\begin{aligned} \|D_j(I - H_j \cdots H_1)\| &= \|D_j(I - H_j) + D_j H_j D_{j-1}(I - H_{j-1} \cdots H_1)\| \\ &\le \|D_j(I - H_j)\| + \|D_{j-1}(I - H_{j-1} \cdots H_1)\|. \end{aligned} \tag{16}$$

Therefore, if we develop a positive sequence $\{\eta_j\}$ such that

$$\eta_0 = 0 \quad \text{and} \quad \eta_j \ge \eta_{j-1} + \|D_j(I - H_j)\|, \quad j = 1, \ldots, m, \tag{17}$$

it follows from (14) and (16) that

$$\|\bar{Z} - Z\| \le \eta_m. \tag{18}$$

The quantity needed to define $\{\eta_j\}$ is an upper bound on $\|D_j(I - H_j)\|$, which we shall obtain by examining the structure of the j-th Householder matrix H_j. In order to simplify this process, we prove the following lemma, which shows that the sequence of Householder matrices needed to triangularize a given matrix is unaffected by postmultiplication by an upper-triangular matrix.

Lemma 1. *Let*

$$H_m \cdots H_1 A = \begin{pmatrix} R \\ 0 \end{pmatrix}$$

represent the reduction to triangular form of the full-rank matrix A by Householder transformations as described in Section 2. Let S be a nonsingular upper-triangular matrix, and let $A' = AS$. If

$$H'_m \cdots H'_1 A' = \begin{pmatrix} R' \\ 0 \end{pmatrix}$$

represents the triangularization of A' by Householder transformations, then

$$H_i = H_i', \quad i = 1, \ldots, m.$$

Proof. The proof is by induction. Let a_i denote the ith column of A, and similarly for a_i' and A'. To begin the induction, note that since a_1' is a nonzero multiple of a_1, it follows that $H_1 = H_1'$. Then suppose, inductively, that $H_i = H_i'$, $i = 1, \ldots, k-1$. At the kth step, H_k is determined by the last $n - k + 1$ components of $H_{k-1} \cdots H_1 a_k$, and likewise H_k' is determined by the last $n - k + 1$ components of $H_{k-1}' \cdots H_1' a_k'$. By definition of A' and our inductive hypothesis, we have

$$\begin{aligned} H_{k-1}' \cdots H_1' a_k' &= H_{k-1}' \cdots H_1'(s_{k,k}a_k + s_{k-1,k}a_{k-1} + \cdots + s_{1,k}a_1) \\ &= H_{k-1} \cdots H_1(s_{k,k}a_k + s_{k-1,k}a_{k-1} + \cdots + s_{1,k}a_1). \end{aligned}$$

By construction of $H_{k-1}, \ldots, H_1$, the last $n - k + 1$ components of $H_{k-1} \cdots H_1 a_i$, $i < k$, are zero. Therefore, the last $n - k + 1$ components of $H_{k-1}' \cdots H_1' a_k'$ are a nonzero multiple of the last $n - k + 1$ components of $H_{k-1} \cdots H_1 a_k$, and it follows that $H_k' = H_k$. ∎

From (12), the matrix to be triangularized is

$$\tilde{R} = \begin{pmatrix} R \\ 0 \end{pmatrix} + E = WR,$$

where

$$W \equiv \begin{pmatrix} I \\ 0 \end{pmatrix} + \Delta \quad \text{and} \quad \Delta \equiv ER^{-1}. \tag{19}$$

Because of Lemma 1, the set of transformations that triangularize $\tilde{R}$ are the same as those that triangularize W (a perturbation of the identity).

Let $\tilde{W}_j$ denote the matrix to be reduced at the j-th step, i.e.,

$$\tilde{W}_j \equiv H_{j-1} \cdots H_1 W,$$

where the first $j - 1$ columns of $\tilde{W}_j$ are already in upper-triangular form. Let $\tilde{w}_j$ denote the j-th column of $\tilde{W}_j$; let t_j denote the first $j - 1$ components of $\tilde{w}_j$, ω_j its j-th component, and $\tilde{v}_j$ its last $n - j$ components, so that

$$\tilde{w}_j^T \equiv (t_j^T, v_j^T) \quad \text{with} \quad v_j^T = (\nu_j, \tilde{v}_j^T). \tag{20}$$

Thus, v_j is the vector to be reduced at step j. From (7), the Householder vector u_j is defined by

$$u_j^T = (\overbrace{0, \ldots, 0}^{j-1}, \nu_j + \operatorname{sign}(\nu_j)\|v_j\|, \tilde{v}_j^T). \tag{21}$$

Note that

$$\sqrt{2}\|v_j\| \le \|u_j\| \le 2\|v_j\|. \tag{22}$$

Using norm inequalities and (22), we obtain

$$\|D_j(I - H_j)\| = \frac{2\|D_j u_j u_j^T\|}{\|u_j\|^2} \le \frac{2\|\tilde{v}_j\|\|u_j\|}{\|u_j\|^2} = 2\,\frac{\|\tilde{v}_j\|}{\|u_j\|} \le \sqrt{2}\,\frac{\|\tilde{v}_j\|}{\|v_j\|}. \tag{23}$$

It follows immediately that the perturbation in components $j+1$ through n of any vector w after application of H_j satisfies:

$$\|D_j(w - H_j w)\| \le \sqrt{2}\,\|w\|\frac{\|\tilde{v}_j\|}{\|v_j\|}. \tag{24}$$

We now develop an upper bound for $\|\tilde{v}_j\|$ and a lower bound for $\|v_j\|$. Let δ_j denote the norm of the j-th column of Δ (which, from (19), is also a bound on the norm of the vector of subdiagonal elements in the j-th column of W). Because Householder matrices are orthogonal, the j-th column $\tilde{w}_j$ of $\tilde{W}_j$ satisfies

$$1 - \delta_j \le \|\tilde{w}_j\| \le 1 + \delta_j. \tag{25}$$

To obtain a lower bound for $\|v_j\|$, we repeatedly apply (24) and (25) to bound the perturbation in components j through n of column j of $\tilde{W}_j$. Formally, let the set of positive values $\{\xi_{j,j}\}$, $j = 1, \ldots, m$ be defined as follows:

$$\begin{aligned} \xi_{j,1} &= 1 - \delta_j \\ \xi_{j,i+1} &= \xi_{j,i} - \delta_i \frac{\sqrt{2}(1+\delta_j)}{\xi_{i,i}}, \quad i = 1, \ldots, j-1. \end{aligned} \tag{26}$$

Then v_j, the vector to be reduced at the j-th step, satisfies

$$\|v_j\| \ge \xi_{j,j}. \tag{27}$$

(We assume that $\|\Delta\|$ is sufficiently small so that $\xi_{j,j}$ remains positive.)

Since $\tilde{v}_j$ corresponds to the subdiagonal elements in column j of $\tilde{W}_j$, we define the sequence $\{\mu_{j,j}\}$, $j = 1, \ldots, m$, as

$$\begin{aligned} \mu_{j,1} &= \delta_j \\ \mu_{j,i+1} &= \mu_{j,i} + \delta_i \frac{\sqrt{2}(1+\delta_j)}{\xi_{i,i}}, \quad i = 1, \ldots, j-1. \end{aligned} \tag{28}$$

It follows from (24) and (25) that

$$\|\tilde{v}_j\| \le \mu_{j,j}. \tag{29}$$

Therefore, using (23), (27) and (29), we have

$$\begin{aligned} \|D_j(I - H_j)\| &\le \sqrt{2}\,\frac{\|\tilde{v}_j\|}{\|v_j\|} \\ &\le \sqrt{2}\,\frac{\mu_{j,j}}{\xi_{j,j}}. \end{aligned} \tag{30}$$

Applying (30) to (17) and (18), we obtain the following bound:

$$\|\bar{Z} - Z\| \le \sqrt{2}\left(\frac{\mu_{1,1}}{\xi_{1,1}} + \cdots + \frac{\mu_{m,m}}{\xi_{m,m}}\right). \tag{31}$$

This expression is intractable as it stands. However, when $\|\Delta\|$ is small, we can obtain a simple expression for the upper bound in (31). First, note that if $\delta_j \ll 1$, then (26) implies that $\xi_{j,j}$ is of order unity for all j. It follows from (28) that the growth in $\mu_{j,j}$ is approximately *linear*, i.e.,

$$\mu_{j,j} \le \sqrt{2}\sum_{i=1}^{j} \delta_j.$$

Let

$$\delta \equiv \|ER^{-1}\|, \tag{32}$$

so that $\delta_j \le \delta$ for all j; then

$$\mu_{j,j} \le j\sqrt{2}\,\delta.$$

Substituting in the bound on $\|\bar{Z} - Z\|$ in (31), we obtain

$$\|\bar{Z} - Z\| \le \frac{\sqrt{2}}{2} m(m+1)\,\delta. \tag{33}$$

Let α denote the smallest singular value of A. Since $\|R^{-1}\| \le \alpha^{-1}$ and $\delta \le \|E\|\|R^{-1}\|$, it follows from (33) that

$$\|\bar{Z} - Z\| \le \rho\alpha^{-1}\|E\|, \tag{34}$$

where ρ depends only on m. As in standard error analysis, we highlight the dependence of the bound on the condition number of A by rewriting (34) as

$$\|\bar{Z} - Z\| \le \rho\,\mathrm{cond}(A)\,\frac{\|E\|}{\|A\|}.$$

where $\mathrm{cond}(A)$ is the ratio of the largest to the smallest singular values of A.

The proof of uniform continuity of Z at $\hat{x}$ is almost immediate. First, note that ρ and α are independent of δx. Second, recall that $\epsilon \to 0$ as $\epsilon_x \to 0$. It follows directly from (34) that

$$\lim_{\|\delta x\| \to 0} \bar{Z} = Z.$$

The bound (34) is interesting because, although Z is not unique, the null space itself (denote by $\mathcal{Q}$) is. Let $\overline{\mathcal{Q}}$ denote the null space of $\bar{A}^T$ from (11). If we measure the distance between $\mathcal{Q}$ an $\overline{\mathcal{Q}}$ by the norm of the difference of the projectors onto them, then $\mathcal{Q}$ and $\overline{\mathcal{Q}}$ differ by a quantit that is asymptotically bounded by $\alpha^{-1}\|E\|$. In Davis and Kahan (1970), it is shown that ther exists a rotation P such that $P\mathcal{Q} = \overline{\mathcal{Q}}$, and $\|I - P\|$ is minimal (in this case, approximatel $\alpha^{-1}\|E\|$). Thus, the choice $\bar{Z} = PZ$ would provide the "best" algorithm for updating Z. Tl bound (34) is larger by a factor of order m^2 than the bound corresponding to the optimal choice of rotation. However, for some matrices E, $\alpha^{-1}\|E\|$ may be a substantial overestimate of $\|ER^{-1}\|$, which may in turn be a substantial overestimate of δ_j for some j. This will be illustrated by Example 2 in Section 5.

3.2. Regularized Householder transformations; perturbation in Q. Although the matrix Z obtained using ordinary Householder matrices undergoes small perturbations in a neighborhood of $\hat{x}$, the same does *not* hold for Y (the first m columns of Q). In fact, the effect of applying each set of transformations $\{H_j\}$ is to change the signs of the columns of Y. Thus, no bound analogous to (34) can be obtained for Y. However, the difficulty can be circumvented by defining the *regularized Householder transformation* $\bar{H}_j$ to be

$$\bar{H}_j = \bar{D}_j H_j, \quad \text{where} \quad \bar{D}_j = \operatorname{diag}(\overbrace{1,\ldots,1}^{j-1}, -1, 1,\ldots,1), \tag{35}$$

i.e., $\bar{H}_j$ is H_j with the sign of its j-th row reversed.

In this section, we derive a bound on $\|\bar{Q} - Q\|$ where $\bar{Q}$ is obtained from Q by the procedure of Section 2, but using *regularized* rather than standard Householder transformations. Because the derivation of the bound is so similar to that for $\|\bar{Z} - Z\|$, we simply highlight the major differences.

The relationship analogous to (16) for regularized Householder transformations is

$$\|I - \bar{H}_j \cdots \bar{H}_1\| \le \|I - \bar{H}_j\| + \|I - \bar{H}_{j-1} \cdots \bar{H}_1\|.$$

Hence, if we derive a sequence $\{\bar{\eta}_j\}$ such that

$$\bar{\eta}_0 = 0, \quad \bar{\eta}_j \ge \bar{\eta}_{j-1} + \|I - \bar{H}_j\|,$$

then

$$\|\bar{Q} - Q\| \le \bar{\eta}_m.$$

Lemma 1 applies to regularized Householder matrices, so that we need to consider only matrices of the form W in (19).

The critical quantity to be determined is a bound on $\|I - \bar{H}_j\|$. To illustrate the process, consider $I - \bar{H}_1 = I - \bar{D}_1 H_1$. Using (20) and (21), v will denote the vector to be reduced, and the corresponding Householder vector u is given by

$$u = \begin{pmatrix} \nu + \operatorname{sign}(\nu)\|v\| \\ \tilde{v} \end{pmatrix}.$$

By definition of $\bar{H}_1$, we have

$$I - \bar{H}_1 = \frac{2}{\|u\|^2} \begin{pmatrix} \|u\|^2 - u_1^2 & -u_1u_2 & \ldots & -u_1u_n \\ u_1u_2 & u_2^2 & \ldots & u_2u_n \\ \vdots & \vdots & \vdots & \vdots \\ u_1u_n & u_2u_n & \ldots & u_n^2 \end{pmatrix}.$$

The Frobenius norm of this matrix may be obtained by direct computation and bounded using (22), giving

$$\|I - \bar{H}_1\|_F = 2\sqrt{2}\,\frac{\|\tilde{v}\|}{\|u\|} \le 2\,\frac{\|\tilde{v}\|}{\|v\|}.$$

Since $\|\cdot\|_2 \le \|\cdot\|_F$, the following lemma follows immediately.

Lemma 2. *Let $\bar{H}_j$ be the regularized Householder transformation defined by (35), (20) and (21); then*

$$\|I - \bar{H}_j\| \le 2\frac{\|\tilde{v}_j\|}{\|v_j\|}. \tag{36}$$

Exactly as for (24) through (31), we can then derive

$$\|\bar{Q} - Q\| \le 2\left(\frac{\mu_{1,1}}{\xi_{1,1}} + \cdots + \frac{\mu_{m,m}}{\xi_{m,m}}\right).$$

(Note that this differs from (31) only in the constant multiplying the right-hand side.) When $\|\Delta\|$ is small, we have the same form of bound as in (34), namely

$$\|\bar{Q} - Q\| \le \rho\alpha^{-1}\|E\|,$$

where ρ depends only on m. Continuity of Q at $\hat{x}$ follows exactly as for Z.

4. Convergence of Z and Q

One reason for interest in the continuity of Z is in proving local convergence results for nonlinearly constrained optimization methods that maintain estimates of the projected Hessian of the Lagrangian function (e.g., Coleman and Conn, 1984a, b). Hence, we now turn to the computation of Z within an iterative method that generates a sequence $\{x_k\}$, where *Q_{k+1} is computed from Q_k using the procedure of Section 2.*

We assume that $\{x_k\}$ converges to a point x^* such that $A(x^*)$ has full rank. Thus, there exists an integer K_1 such that for all $k > K_1$, $A(x_k)$ has full rank; we shall consider only such values of k. We further assume that

$$\sum_{k=0}^{\infty} \|x_k - x^*\| < +\infty.$$

This implies that for any $\epsilon > 0$, there exists an integer K such that for all $l > k > K \ge K_1$,

$$\|x_l - x_{l-1}\| + \ldots + \|x_{k+1} - x_k\| < \epsilon. \tag{37}$$

For a given value of ϵ in (37), we shall consider only values of the iteration count that exceed the associated K.

The bound (34) shows that, for sufficiently large K, there exists a positive constant M, *independent of k*, such that for all $k > K$,

$$\|Z_{k+1} - Z_k\| < M\|x_{k+1} - x_k\|.$$

Therefore, we have for all $l > k > K$,

$$\begin{aligned} \|Z_l - Z_k\| &\le \|Z_l - Z_{l-1}\| + \ldots + \|Z_{k+1} - Z_k\| \\ &< M(\|x_l - x_{l-1}\| + \ldots + \|x_{k+1} - x_k\|). \end{aligned} \tag{38}$$

Because of (37), the sum on the right in (38) can be made as small as desired by appropriate choice of K, and hence $\|Z_l - Z_k\|$ can be made as small as desired. Thus $\{Z_k\}$ is a Cauchy sequence, and therefore converges to a limit Z^* as $\{x_k\}$ converges to x^*. We emphasize that *the limit Z^* depends on the sequence $\{x_k\}$.*

If regularized Householder matrices are used to define Q_{k+1} from Q_k, exactly the same result holds for the full matrix Q.

5. Numerical examples

In this section, we illustrate some properties of the method with two simple examples.

Example 1. Let $x_0 = (1,0,1)^T$ and $x^* = (0,1,2)^T$. We define the function $a(x) = x$, and consider the following sequence, which begins at x_0 and converges to x^* so as to satisfy (37):

$$x_k = \begin{pmatrix} (-1/2)^k \\ 1-(-1/2)^k \\ 2-1/2^k \end{pmatrix}.$$

With $Q_{-1} \equiv I$, the matrix Q_0 is the Householder matrix

$$Q_0 = \begin{pmatrix} -.70711 & 0 & -.70711 \\ 0 & 1 & 0 \\ -.70711 & 0 & .70711 \end{pmatrix}.$$

For each k, Z_k is the last two columns of Q_k.

All computation was performed using double-precision arithmetic on an IBM 3081, corresponding to about 16 decimal digits of precision. All numbers shown are rounded to five figures. At steps 10 and 11, where $\|x_{11} - x_{10}\| = 2.1284 \times 10^{-3}$, we have

$$Q_{10} = \begin{pmatrix} -.000437 & -.28042 & -.95988 \\ -.89451 & .85868 & -.25066 \\ -.44704 & -.42900 & .12574 \end{pmatrix},$$

$$Q_{11} = \begin{pmatrix} -.000218 & -.28042 & -.95988 \\ .89430 & .85839 & -.25088 \\ .44748 & -.42958 & .12530 \end{pmatrix},$$

which corresponds to $\|Z_{11} - Z_{10}\|_F = 8.1736 \times 10^{-4}$. Note the change of sign in the first column of Q.

A sequence of orthogonal matrices was similarly generated for the following sequence $\{y_k\}$, which also begins at x_0 and converges to x^*:

$$y_k = \begin{pmatrix} 1/2^k \\ 1 - 1/2^k \\ 2 - 1/2^k \end{pmatrix}.$$

Note that $Q(y_0) = Q(x_0)$. However, all subsequent matrices differ for the two sequences.

In both cases, the Z matrices converge, but to different limits:

$$\lim_{k\to\infty} Z(x_k) = \begin{pmatrix} .28042 & -.95988 \\ .85854 & -.25082 \\ -.42927 & .12541 \end{pmatrix},$$

$$\lim_{k\to\infty} Z(y_k) = \begin{pmatrix} -.19371 & -.98106 \\ .87749 & -.17326 \\ -.43874 & .08663 \end{pmatrix}.$$

The first column of the limiting matrix $Q(x^*)$ is $\pm(0,\ .89443,\ .44721)^T$.

For comparison, the Q matrix that would result from applying a standard Householder reduction at x^* is given by

$$Q(x^*) = \begin{pmatrix} 0 & -.89443 & -44721 \\ -.89443 & -.4 & .8 \\ -.44721 & .8 & -.4 \end{pmatrix}.$$

Example 2. The second example shows how the relationship between E and R in (19) can affect the actual change in Z, although the bound (34) remains unchanged. Let A be 6×3, with R given by

$$R = \begin{pmatrix} 10^7 & 1 & 1 \\ & 10 & 1 \\ & & 1 \end{pmatrix}.$$

The smallest singular value of R is of order unity. Consider a matrix E such that $\|E\| = O(1)$; then (34) implies that the change in Z can be of order unity.

This bound is achieved, for example, if E is given by

$$E = \begin{pmatrix} 10^{-7} & 10^{-7} & 1 \\ \vdots & \vdots & \vdots \\ 10^{-7} & 10^{-7} & 1 \end{pmatrix},$$

since $\|ER^{-1}\|$ is of order unity, i.e., similar in magnitude to $\|E\|\|R^{-1}\|$. However, if

$$E = \begin{pmatrix} 1 & 10^{-7} & 10^{-7} \\ \vdots & \vdots & \vdots \\ 1 & 10^{-7} & 10^{-7} \end{pmatrix},$$

then the perturbation in Z is of order 10^{-7} (much less than the bound) because $\|ER^{-1}\| = O(10^{-7}) \ll \|E\|\|R^{-1}\|$.

6. Comparison with alternative procedures

In this section, we compare the procedure of Section 2 with an alternative technique for obtaining an explicit matrix Z. We emphasize that *an explicit Q is required* in the most popular algorithms today for solving constrained problems with *inequality* constraints. An implicit Q is suitable if A changes only by the addition of columns. However, any other change to A can be made efficiently only by access to an explicit Q. Inequality constraints are most often treated by posing a quadratic programming (QP) subproblem with linearized (inequality) versions of the original constraints (see Powell, 1983, for a survey of sequential quadratic programming methods). The QP is solved by developing a *working set A* that undergoes the addition *and deletion* of columns until it becomes the active set of the QP. Furthermore, if simple bound constraints are treated separately from general linear constraints, the matrix A is also subject to the addition and/or deletion of *rows* (see Gill *et al.*, 1984a, for details of the update procedures).

The most obvious alternative to the method of Section 2 is to apply a standard Householder procedure in which the Householder vectors are stored in compact form during the triangularization; we shall refer to this as the *implicit* procedure. Assuming that the m_L transformations corresponding to constant columns of A are retained, the matrices $\bar{R}$ and $\bar{Q}$ of Section 2 can be computed using the standard Householder procedure in $m_N(2nm_L - m_L^2)$ operations to apply the m_L fixed transformations to A_N, and $\frac{2}{3}m_N^3 + m_N^2(n-m)$ operations to produce the desired triangular form. The explicit matrix Q is then formed by multiplying the transformations together in reverse order, which requires $2nm(n-m) + \frac{2}{3}m^3$ operations.

When no linear constraints are present ($m_L = 0$), the implicit procedure requires less storage and work than the explicit procedure. However, as the proportion of linear constraints increases, the explicit procedure eventually requires *less work* (in effect, because the implicit procedure must repeatedly multiply together the Householder transformations corresponding to linear constraints in order to obtain the explicit matrix Q). We stress this point because many optimization problems contain a significant proportion of linear constraints. Although it is simpler to treat all constraints as nonlinear for expository purposes (as we have done in Section 3), their existence should be considered when analyzing the work associated with a practical algorithm.

To summarize, the procedure of Section 2 ensures the continuity properties of Z needed in many constrained optimization algorithms, and can easily be extended to imply continuity of Q. Furthermore, its cost is comparable to (or even less than) that of the implicit procedure when the problem contains a significant proportion of linear constraints.

References

Byrd, R. H. and Schnabel, R. B. (1984). Continuity of the null space basis and constrained optimization, Report CU-CS-272-84, Department of Computer Science, University of Colorado, Boulder, Colorado.

Coleman, T. F. and Conn, A. R. (1982). Nonlinear programming via an exact penalty function: asymptotic analysis, *Mathematical Programming* **24**, pp. 123–136.

Coleman, T. F. and Conn, A. R. (1984). On the local convergence of a quasi-Newton method for the nonlinear programming problem, *SIAM Journal on Numerical Analysis* **21**, pp. 755–769.

Coleman, T. F. and Sorensen, D. C. (1984). A note on the computation of an orthogonal basis for the null space of a matrix, *Mathematical Programming* **29**, pp. 234–242.

Davis, C. and Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation, III, *SIAM Journal on Numerical Analysis* **7**, pp. 1–46.

Gill, P. E., Murray, W., Saunders, M. A. and Wright, M. H. (1983). On the representation of a basis for the null space, Report SOL 83-19, Department of Operations Research, Stanford University, Stanford, California.

Gill, P. E., Murray, W., Saunders, M. A. and Wright, M. H. (1984a). Procedures for optimization problems with a mixture of bounds and general linear constraints, *ACM Transactions on Mathematical Software* **10**, pp. 282–298.

Gill, P. E., Murray, W., Saunders, M. A. and Wright, M. H. (1984b). User's guide to SOL/QPSOL (Version 3.2), Report SOL 84-6, Department of Operations Research, Stanford University, Stanford, California.

Gill, P. E., Murray, W., Saunders, M. A. and Wright, M. H. (1984c). User's guide to SOL/NPSOL (Version 2.1), Report SOL 84-71, Department of Operations Research, Stanford University, Stanford, California.

Powell, M. J. D. (1983). "Variable metric methods for constrained optimization", in *Mathematical Programming: The State of the Art* (A. Bachem, M. Grötschel and B. Korte, eds.), pp. 288–311, Springer-Verlag, Berlin.

Stewart, G. W. (1973). *Introduction to Matrix Computations*, Academic Press, London and New York.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 21592.5-MA	2. GOVT ACCESSION NO. AD-A157681	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PROPERTIES OF A REPRESENTATION OF A BASIS FOR THE NULL SPACE		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Philip E. Gill, Walter Murray, Michael A. Saunders, G.W. Stewart, and Margaret H. Wright		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0267 DAAG29-84-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Operations Research - SOL Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-047-143
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research - Dept. of the Navy 800 N. Quincy Street Arlington, VA 22217		12. REPORT DATE February 1985
		13. NUMBER OF PAGES 14
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

constrained optimization — null space basis

QR factorization — regularized Householder transformations

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(See next page)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SOL 85-1: PROPERTIES OF A REPRESENTATION OF A BASIS FOR THE NULL SPACE by Philip E. Gill, Walter Murray, Michael A. Saunders, G. W. Stewart and Margaret H. Wright.

Given a rectangular matrix $A(x)$ that depends on the independent variables x, many constrained optimization methods involve computations with $Z(x)$, a matrix whose columns form a basis for the null space of $A^T(x)$. When A is evaluated at a given point, it is well known that a suitable Z (satisfying $A^T Z = 0$) can be obtained from standard matrix factorizations. However, Coleman and Sorensen have recently shown that standard orthogonal factorization methods may produce orthogonal bases that do not vary continuously with x; they also suggest several techniques for adapting these schemes so as to ensure continuity of Z in the neighborhood of a given point.

This paper is an extension of an earlier note that defines the procedure for computing Z. Here, we first describe how Z can be obtained by updating an explicit QR factorization with Householder transformations. The properties of this representation of Z with respect to perturbations in A are discussed, including explicit bounds on the change in Z. We then introduce regularized Householder transformations, and show that their use implies continuity of the full matrix Q. The convergence of Z and Q under appropriate assumptions is then proved. Finally, we indicate why the chosen form of Z is convenient in certain methods for nonlinearly constrained optimization.

END

FILMED

9-85

DTIC